Derivative-Free Estimation of the Score Vector and Observed Information Matrix with Application to State-Space Models

Arnaud Doucet[1], Pierre E. Jacob[2] and Sylvain Rubenthaler[3]

[1]University of Oxford, United Kingdom.
[2]National University of Singapore.
[3]Université de Nice-Sophia Antipolis, France.

May 1, 2013

Abstract

Ionides et al. [12, 13] have recently introduced an original approach to perform maximum likelihood parameter estimation in state-space models which only requires being able to simulate the latent Markov model according to its prior distribution. Their methodology relies on an approximation of the score vector for general statistical models based upon an artificial posterior distribution and bypasses the calculation of any derivative. Building upon this insightful work, we provide here a simple "derivative-free" estimator of the observed information matrix based upon this very artificial posterior distribution. However for state-space models where sequential Monte Carlo computation is required, these estimators have too high a variance and need to be modified. In this specific context, we derive new derivative-free estimators of the score vector and observed information matrix which are computed using sequential Monte Carlo approximations of smoothed additive functionals associated with a modified version of the original state-space model.

Keywords: Maximum likelihood, Score vector, Observed information matrix, Sequential Monte Carlo, Smoothing, State-space models.

1 Introduction

Consider a random variable Y taking values in a measurable space $\mathcal{Y}$. Given $\theta \in \mathbb{R}^d$, we assume that Y follows a probability density function $p_Y(y; \theta)$ w.r.t. a σ-finite dominating measure denoted $\upsilon(dy)$. Given $Y = y$, the likelihood is denoted by $\mathcal{L}(\theta) = p_Y(y; \theta)$ and the log-likelihood by $\ell(\theta) = \log p_Y(y; \theta)$. Assuming that $\ell(\theta)$ is twice differentiable, we are interested in calculating the score vector $\ell^{(1)}(\theta)$ and the observed information matrix $-\ell^{(2)}(\theta)$ whose r^{th} component $\ell_r^{(1)}$ and $(r, s)^{\text{th}}$ component $-\ell_{r,s}^{(2)}(\theta)$ are given for $r, s = 1, \ldots, d$ by

$$\ell_r^{(1)} = \frac{\partial \ell(\theta)}{\partial \theta^r} \quad \text{and} \quad -\ell_{r,s}^{(2)}(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta^r \partial \theta^s}. \tag{1}$$

The score vector and observed information matrix are useful both algorithmically and statistically. Algorithmically, they can be used to build efficient maximum likelihood estimation techniques as in [12, 13] or to build efficient MCMC proposals relying on the local geometry of the target distribution [10]. Statistically, the observed information matrix can be used to estimate the variance of the maximum likelihood estimate.

Exact calculations of the score vector and observed information matrix are only possible for models where $\ell(\theta)$ can be evaluated exactly. For complex latent variable models, these quantities are typically computed using Monte Carlo approximations of the Fisher and Louis identities [4, 17]. However there are many important scenarios where this is not even a viable option. For example for numerous state-space models arising in applied science, we are only able to obtain sample paths from the latent Markov process but we have access to the expression of neither its transition kernel nor its derivatives [12, 13]. This prohibits the numerical implementation of the Fisher and Louis identities. It is thus useful to develop a simple method to obtain estimates of the score vector and observed information matrix which, beyond the specification of the statistical model, requires a minimum amount of input from the user but can outperform finite difference approximations.

For the score vector, such a method has been recently proposed in [12, 13]. The main idea of the authors is to introduce an artificial random parameter $\widetilde{\Theta}$ with prior centred around θ. They establish that the expectation

1

of $\widetilde{\Theta} - \theta$ w.r.t. the posterior associated to this prior and the likelihood $\mathcal{L}(\theta)$ has components approximately proportional to the components of $\ell^{(1)}(\theta)$; the approximation improving as the artificial prior shrinks around θ. In a state-space context where sequential Monte Carlo approximations are required, the direct application of this idea provides a high variance estimator. The authors propose a lower variance estimator which is computed using the optimal filter associated to a modified version of the original state-space model where an artificial random walk dynamics initialized at the parameter θ is introduced.

In this paper, our contributions are two-fold. First, in Section 2, we extend the idea in [12, 13] for approximating the score vector to the approximation of the observed information matrix. We show that this latter is directly related to the covariance of the artificial posterior associated to $\widetilde{\Theta}$ when the prior is carefully selected. Additionally we sharpen the theoretical results provided in [13] and use these results to compare the proposed estimators to finite difference estimators in terms of optimal rates of convergence of the mean squared error. These results hold for general statistical models. Second, in the specific context of state-space models, we propose in Section 3 original estimators of the score vector and observed information matrix. These are computed using the optimal smoother associated to a modified state-space model which enjoys nicer theoretical properties than the one considered in [13]. This allows us to obtain quantitative bounds for the sequential Monte Carlo implementations of the estimators.

All proofs are postponed to the appendix.

2 Derivative-free estimates of the score vector and observed information matrix

2.1 An artificial Bayesian model

We follow here the approach initiated in [13] and introduce a stochastically perturbed version of the original model corresponding to a pair of random variables $(\widetilde{\Theta}, \widetilde{Y})$ having a joint probability density on $\mathbb{R}^d \times \mathcal{Y}$

$$\widetilde{p}_{\widetilde{\Theta}, \widetilde{Y}}\left(\widetilde{\theta}, y; \theta, \tau\right) = \tau^{-d} \kappa \left\{\tau^{-1}\left(\widetilde{\theta} - \theta\right)\right\} p_Y\left(y; \widetilde{\theta}\right) \tag{2}$$

where $\tau > 0$ is a scale parameter and $\kappa(\cdot)$ a probability density on $\mathbb{R}^d$.

Our main result in this section relates the expectation of an arbitrary function $h(\widetilde{\Theta} - \theta)$ w.r.t. the posterior $\widetilde{p}_{\widetilde{\Theta}|\widetilde{Y}}(\theta \mid y; \theta, \tau)$ of $\widetilde{\Theta}$ given $\widetilde{Y} = y$ defined through equation (2) to the entries of the score vector and observed information matrix given in (1). We will denote by $\widetilde{\mathbb{E}}_{\theta_{1:T}, \tau}$ the expectation w.r.t. this artificial posterior. To present this result, we need to introduce some additional notation. If $f : \mathbb{R}^d \to \mathbb{R}$ is enough times differentiable, its kth-order differential at θ is a k-linear application from $\mathbb{R}^{d \times k}$ to $\mathbb{R}$ denoted $f^{(k)}(\theta)$ and we write

$$f^{(k)}(\theta).u^{\otimes k} = \sum_{1 \leq i_1, \ldots, i_k \leq d} \frac{\partial^k f(\theta)}{\partial \theta_{i_1} \ldots \partial \theta_{i_k}} u_{i_1} \ldots u_{i_k} \tag{3}$$

where u_i denotes the i-th component of u. For any vector $u \in \mathbb{R}^d$ and matrix $v \in \mathbb{R}^{d \times d}$ we denote by $|u|$ and $|v|$ the L_1-norm: $|u| = \sum_{i=1}^d |u_i|$ and $|v| = \sum_{i=1}^d \sum_{j=1}^d |v_{ij}|$. For a vector-valued function $f = (f_1, \ldots, f_m)^{\mathsf{T}}$, we write $\int f(u) \, du$ for the vector $\left(\int f_1(u) \, du, \ldots, \int f_m(u) \, du\right)^T$.

Our results rely on the following assumptions.

Assumption 1 κ is a symmetric probability density on $\mathbb{R}^d$ w.r.t. Lebesgue measure. For any $k \geq 1$, $1 \leq i_1, \ldots, i_k \leq d$ and $\beta_1, \ldots, \beta_k \geq 1$ there exists $C(i_1, \ldots, i_k, \beta_1, \ldots, \beta_k) < \infty$ such that

$$\int \left| u_{i_1}^{\beta_1} u_{i_2}^{\beta_2} \cdots u_{i_k}^{\beta_k} \right| \kappa(u) du \leq C(i_1, \ldots, i_k, \beta_1, \ldots, \beta_k)$$

and the covariance matrix $\Sigma = (\sigma_{i,j})_{i,j=1}^d$ associated to κ is non-singular so $\sigma_i^2 := \sigma_{i,i} > 0$.

Assumption 2 κ is such that

$$\exists \gamma, \delta, M > 0, \quad \forall u \in \mathbb{R}^d \quad |u| > M \implies \kappa(u) < e^{-\gamma|u|^\delta}.$$

Assumption 3 *The log-likelihood function $\ell : \mathbb{R}^d \to \mathbb{R}$ is four times continuously differentiable and, for δ defined as in Assumption 2, the associated likelihood $\mathcal{L} : \mathbb{R}^d \to \mathbb{R}$ satisfies:*

$$\forall \theta \in \mathbb{R}^d \quad \exists\, 0 < \eta < \delta \quad \exists\, \epsilon, D > 0 \quad \forall u \in \mathbb{R}^d \quad \mathcal{L}(\theta + u) \leq D e^{\epsilon |u|^\eta}.$$

Assumptions 2 and 3 ensures that if the likelihood function is not bounded then it goes to infinity slowly enough so that the prior κ "compensates". These assumptions are not restrictive given κ can be selected by the user; see however the comment after Assumption 4. The following result then holds.

Theorem 1 *Suppose Assumptions 1-2-3. For any $\theta \in \mathbb{R}^d$, and $h : \mathbb{R}^d \to \mathbb{R}^m$ satisfying*

$$|h(u)| \leq c\,|u|^\alpha \tag{4}$$

for some constants $\alpha \geq 0$ and $c > 0$, we have

$$
\begin{aligned}
\widetilde{\mathbb{E}}_{\theta,\tau}\left\{ h(\widetilde{\Theta} - \theta) \middle|\, \widetilde{Y} = y \right\} &= \int h(\tau u)\kappa(u)du + \tau \int h(\tau u)\,\ell^{(1)}(\theta).u\,\kappa(u)du \\
&\quad + \frac{\tau^2}{2}\int\left\{ h(\tau u) - \int h(\tau u)\kappa(u)du \right\}\left\{ \ell^{(2)}(\theta).u^{\otimes 2} + (\ell^{(1)}(\theta).u)^2 \right\}\kappa(u)du \\
&\quad + \tau^3 \int h(\tau u)\left[\frac{1}{3!}\left\{\ell^{(1)}(\theta).u\right\}^3 + \frac{1}{2}\left\{\ell^{(1)}(\theta).u\right\}\left\{\ell^{(2)}(\theta).u^{\otimes 2}\right\} + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right]\kappa(u)du \\
&\quad - \frac{\tau^3}{2}\left[\int\left\{\ell^{(2)}(\theta).u^{\otimes 2} + (\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du\right]\left\{\int h(\tau u)\ell^{(1)}(\theta).u\,\kappa(u)du\right\} \\
&\quad + \mathcal{O}(\tau^{4+\alpha}).
\end{aligned}
$$

All our results are presented as asymptotic expansions for each θ; we could have also presented them as uniform upper bounds on the remainder term, for any $\theta \in K$ in some compact set K as in [13].

2.2 Approximation of the score vector and observed information matrix

We detail here two useful consequences of Theorem 1. The first result is a strengthened version of the main result provided in [13] which shows that the rescaled components of $\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta} - \theta \middle|\, \widetilde{Y} = y\right)$ are approximately proportional to the score vector. This is established by applying Theorem 1 to the function $h(u) = u$.

Theorem 2 *Suppose Assumptions 1-2-3. For any $\theta \in \mathbb{R}^d$, there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \leq \eta$*

$$|\ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta} - \theta \middle|\, \widetilde{Y} = y\right)| \leq C\,\tau^2. \tag{5}$$

Whereas the upper bound on the r.h.s of (5) provided in [13] is of order τ, Theorem 1 shows that it is actually of order τ^2. This sharper bound is crucial when comparing theoretically this estimator of the score to finite differences (see Section 2.4). It is additionally possible to approximate the observed information matrix by rescaling the elements of posterior covariance $\widetilde{\mathbb{C}ov}_{\theta,\tau}(\widetilde{\Theta}|\widetilde{Y} = y)$. However, this requires an additional assumption on the artificial prior κ that is for example verified when κ is a multivariate normal with diagonal covariance matrix.

Assumption 4 *κ satisfies $\kappa(u) = \prod_{i=1}^{d} \kappa_i(u_i)$ and is mesokurtic, that is*

$$\Lambda_i := \int u_i^4 \kappa(u)du = 3\sigma_i^4.$$

Note that choosing a multivariate normal distribution for κ in order to satisfy Assumption 4 makes the constraints on the likelihood brought by Assumption 3 more explicit. In this context, we obtain the following result by applying Theorem 1 to $h(u) = uu^{\mathrm{T}}$.

Theorem 3 *Suppose Assumptions 1-2-3-4. For any $\theta \in \mathbb{R}^d$, there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \leq \eta$*

$$\left| -\ell^{(2)}(\theta) + \tau^{-4}\Sigma^{-1}\left\{\widetilde{\mathbb{C}ov}_{\theta,\tau}\left(\widetilde{\Theta}\middle|\, \widetilde{Y} = y\right) - \tau^2\Sigma\right\}\Sigma^{-1}\right| \leq C\tau^2.$$

2.3 Latent variable models and Monte Carlo estimates

The approximations of $\ell^{(1)}(\theta)$ and $\ell^{(2)}(\theta)$ presented in Theorems 2 and 3 will be primarily useful for latent variable models where we have random variables (X, Y) taking values in a measurable space $\mathcal{X} \times \mathcal{Y}$ and following a probability density function

$$p_{X,Y}(x, y; \theta) = p_X(x; \theta) \, p_{Y|X}(y|x; \theta) \tag{6}$$

w.r.t. a σ-finite product dominating measure denoted $\lambda(dx) \, \upsilon(dy)$ that is parameterised by $\theta \in \mathbb{R}^d$. Here X is a latent variable and, given $Y = y$, we have $\ell(\theta) = \log p_Y(y; \theta)$ where $p_Y(y; \theta) = \int p_{X,Y}(x, y; \theta) \, dx$.

In this context, the artificial Bayesian model corresponds to a triplet of random variables $\left(\widetilde{\Theta}, \widetilde{X}, \widetilde{Y} \right)$ having a joint probability density on $\mathbb{R}^d \times \mathcal{X} \times \mathcal{Y}$

$$\widetilde{p}_{\widetilde{\Theta}, \widetilde{X}, \widetilde{Y}} \left(\widetilde{\theta}, x, y; \theta, \tau \right) = \tau^{-d} \kappa \left\{ \tau^{-1} \left(\widetilde{\theta} - \theta \right) \right\} p_{X,Y} \left(x, y; \widetilde{\theta} \right).$$

Theorems 2 and 3 still obviously hold and $\ell^{(1)}(\theta)$ and $\ell^{(2)}(\theta)$ can be estimated by performing a Monte Carlo approximation of the posterior $\widetilde{p}_{\widetilde{\Theta}, \widetilde{X}|\widetilde{Y}}(\widetilde{\theta}, \widetilde{x}|y; \theta, \tau)$, hence of its marginal $\widetilde{p}_{\widetilde{\Theta}|\widetilde{Y}}(\widetilde{\theta}|y; \theta, \tau)$, and then estimating the associated posterior mean $\widetilde{\mathbb{E}}_{\theta, \tau}(\widetilde{\Theta} \mid \widetilde{Y} = y)$ and covariance $\widetilde{\mathbb{C}ov}_{\theta, \tau}(\widetilde{\Theta} \mid \widetilde{Y} = y)$.

2.4 Comparison with finite difference schemes

An alternative "derivative-free" approach to compute $\ell^{(1)}(\theta)$ and $\ell^{(2)}(\theta)$ consists of using finite difference schemes combined to Monte Carlo estimates of $\ell(\theta)$; see for example [2]. For sake of simplicity, consider the case where $\theta \in \mathbb{R}$.

The central finite difference estimator of $\ell^{(1)}(\theta)$ and second central finite difference estimator of $\ell^{(2)}(\theta)$ are given by

$$\widehat{\ell}_{N,h}^{(1)}(\theta) = \frac{\widehat{\ell}_N(\theta + h) - \widehat{\ell}_N(\theta - h)}{2h}, \quad \widehat{\ell}_{N,h}^{(2)}(\theta) = \frac{\widehat{\ell}_N(\theta + h) - 2\widehat{\ell}_N(\theta) + \widehat{\ell}_N(\theta - h)}{h^2} \tag{7}$$

where $\widehat{\ell}_N(\theta + h)$, $\widehat{\ell}_N(\theta - h)$ and $\widehat{\ell}_N(\theta)$ are independent Monte Carlo estimates using N samples. In most applications, these three estimators have both a bias and a variance of order N^{-1}. It can then be shown under mild additional regularity assumptions that the optimal rates of convergence of the mean squared error for $\widehat{\ell}_{N,h}^{(1)}(\theta)$ and $\widehat{\ell}_{N,h}^{(2)}(\theta)$ are

$$\mathbb{E}\left\{ \widehat{\ell}_{N,h}^{(1)}(\theta) - \ell^{(1)}(\theta) \right\}^2 \sim N^{-2/3} \text{ and } \mathbb{E}\left\{ \widehat{\ell}_{N,h}^{(2)}(\theta) - \ell^{(2)}(\theta) \right\}^2 \sim N^{-1/2}$$

for $h \sim N^{-1/6}$ and $h \sim N^{-1/8}$; see [2], Chapter 7, Section 1 for details. Results provided in Appendix B indicate that the optimal rates of convergence of the mean squared error of the estimators

$$\widehat{\ell}_{N,\tau}^{(1)}(\theta) = \tau^{-2} \Sigma^{-1} \left\{ \widehat{\mu}_{N,\tau}(\theta) - \theta \right\}, \quad \widehat{\ell}_{N,\tau}^{(2)}(\theta) = \tau^{-4} \Sigma^{-1} \left\{ \widehat{v}_{N,\tau}(\theta) - \tau^2 \Sigma \right\} \Sigma^{-1} \tag{8}$$

with $\widehat{\mu}_{N,\tau}(\theta)$ and $\widehat{v}_{N,\tau}(\theta)$ being importance sampling estimators of $\widetilde{\mathbb{E}}_{\theta, \tau}(\widetilde{\Theta}|\widetilde{Y} = y)$ and $\widetilde{\mathbb{C}ov}_{\theta, \tau}(\widetilde{\Theta}|\widetilde{Y} = y)$ are similar to the finite difference optimal rates, provided $\tau \sim N^{-1/6}$ and $\tau \sim N^{-1/8}$ respectively.

As pointed out in [2], these results are "rather academic" as the constants in front of the optimals h or τ depend on unknown parameters. Moreover, as observed experimentally in [13] in the context of state-space models, $\widehat{\ell}_{N,\tau}^{(1)}(\theta)$ can outperform significantly $\widehat{\ell}_{N,h}^{(1)}(\theta)$ as $\widehat{\ell}_{N,h}^{(1)}(\theta)$ involves running two independent sequential Monte Carlo filters providing a high variance estimate of the numerator of (7) whereas variance reduction techniques positively correlating $\widehat{\ell}_N(\theta + h)$ and $\widehat{\ell}_N(\theta - h)$ proposed in [15] are not applicable in this model-free context.

3 Estimation of the score vector and observed information matrix for state-space models

3.1 State-space models

Let $\{X_t, Y_t\}_{t \in \mathbb{N}}$ be a stochastic process such that (X_t, Y_t) takes values in a measurable space $\mathcal{X} \times \mathcal{Y}$. For any sequence $\{z_k\}$, let $z_{i:j}$ denote $(z_i, z_{i+1}, ..., z_j)$. The model is specified as follows: $\{X_t\}_{t \in \mathbb{N}}$ is a latent Markov

process of initial density $\nu(x;\theta)$ and homogeneous Markov transition density $f(x|x';\theta)$ w.r.t. a dominating measure $\lambda(dx)$ whereas the observations $\{Y_t\}_{t\in\mathbb{N}}$ are assumed to be conditionally independent given $\{X_t\}_{t\in\mathbb{N}}$ of conditional density $g(y_t|x_t;\theta)$ w.r.t. a dominating measure $\upsilon(dy)$; that is $X_1 \sim \mu(\cdot;\theta)$ and for $t=1,2,\ldots$.

$$X_{t+1}|(X_t = x) \sim f(\cdot|x_{t-1};\theta), \qquad Y_t|(X_t = x) \sim g(\cdot|x_t;\theta). \tag{9}$$

It follows that the joint density of $(X_{1:T}, Y_{1:T})$ is given by

$$p_{X_{1:T},Y_{1:T}}(x_{1:T}, y_{1:T};\theta) = \nu(x_1;\theta)\prod_{t=2}^{T} f(x_t|x_{t-1};\theta)\prod_{t=1}^{T} g(y_t|x_t;\theta). \tag{10}$$

For a realization $Y_{1:T} = y_{1:T}$ of the observations, the log-likelihood of function satisfies

$$\ell(\theta) = \log\, p_{Y_{1:T}}(y_{1:T};\theta), \text{where } p_{Y_{1:T}}(y_{1:T};\theta) = \int p_{X_{1:T},Y_{1:T}}(x_{1:T}, y_{1:T};\theta)\,\lambda(dx_{1:T}). \tag{11}$$

This model is just a specific latent variable model as discussed in Section 2.3 with $X = X_{1:T}$ and $Y = Y_{1:T}$. Hence it is possible to approximate $\ell^{(1)}(\theta)$ and $\ell^{(2)}(\theta)$ using Theorem 2 and Theorem 3 and by computing a Monte Carlo approximation of $\widetilde{p}_{\widetilde{\Theta},\widetilde{X}_{1:T}|Y_{1:T}}(\widetilde{\theta}, x_{1:T}|y_{1:T};\theta,\tau)$. Sequential Monte Carlo methods are the tools of choice to approximate posterior distributions for state-space models. Unfortunately, it is well-documented that standard sequential Monte Carlo methods would provide very high variance estimators of $\widetilde{p}_{\widetilde{\Theta}|Y_{1:T}}(\theta|y_{1:T};\theta,\tau)$ in this context as the parameter $\widetilde{\Theta}$ is static, see [4, 8]. Recently particle Markov chain Monte Carlo [1] and Sequential Monte Carlo squared [5] algorithms have been developed to address such problems and could be used to sample from $\widetilde{p}_{\widetilde{\Theta}|Y_{1:T}}(\widetilde{\theta}|y_{1:T};\theta,\tau)$. We follow here an alternative approach initiated in [13], leading to a natural extension of the approximations described in Section 2.2 and their Monte Carlo estimates described in Section 2.3.

3.2 An artificial Bayesian model

We first extend the model by allowing the parameter θ to change at each time point. We therefore introduce $\theta_{1:T} \in \mathbb{R}^{dT}$ and an extended model such that $\bar{X}_{1:T}, \bar{Y}_{1:T}$ have a joint density defined by

$$\bar{p}_{\bar{X}_{1:T},\bar{Y}_{1:T}}(x_{1:T}, y_{1:T}|\theta_{1:T}) = \nu(x_1;\theta_1)\prod_{t=2}^{T} f(x_t|x_{t-1};\theta_t)\prod_{t=1}^{T} g(y_t|x_t;\theta_t) \tag{12}$$

and we denote the associated log-likelihood of the observations $\bar{Y}_{1:T} = y_{1:T}$ by

$$\bar{\ell}(\theta_{1:T}) = \log\, p_{\bar{Y}_{1:T}}(y_{1:T};\theta_{1:T})$$

We write $\theta^{[T]}$ the vector of $\mathbb{R}^{dT}$ made of T copies of θ concatenated in a column vector. Similarly to Section 2.1, we now introduce artificial random variables $\widetilde{\Theta}_{1:T}$ with prior $\overline{\kappa}$ centred around $\theta_{1:T} \in \mathbb{R}^{dT}$. Given a prior κ on $\mathbb{R}^d$ as in Section 2.1, satisfying Assumptions 1 and 2, we define $\overline{\kappa}$ as

$$\overline{\kappa}(\widetilde{\theta}_{1:T};\theta_{1:T},\tau) = \prod_{t=1}^{T} \tau^{-d}\kappa\left\{\tau^{-1}\left(\widetilde{\theta}_t - \theta_t\right)\right\} \tag{13}$$

and denote by $\bar{\Sigma}_T$ its associated block diagonal covariance matrix. The joint probability density of $(\widetilde{\Theta}_{1:T}, \widetilde{X}_{1:T}, \widetilde{Y}_{1:T})$ is then defined as

$$\widetilde{p}_{\widetilde{\Theta}_{1:T},\widetilde{X}_{1:T},\widetilde{Y}_{1:T}}(\widetilde{\theta}_{1:T}, x_{1:T}, y_{1:T}|\theta_{1:T},\tau) = \bar{p}_{\bar{X}_{1:T},\bar{Y}_{1:T}}(x_{1:T}, y_{1:T}|\widetilde{\theta}_{1:T})\overline{\kappa}(\widetilde{\theta}_{1:T};\theta_{1:T},\tau) \tag{14}$$

and we denote by $\widetilde{\mathbb{E}}_{\theta_{1:T},\tau}$ the expectation with respect to the associated posterior $\widetilde{p}_{\widetilde{\Theta}_{1:T}|\widetilde{Y}_{1:T}}(\widetilde{\theta}_{1:T}|y_{1:T};\theta_{1:T},\tau)$.

In order to apply the results of Section 2.2, we further assume that the log-likelihood $\bar{\ell}(\theta_{1:T})$ associated to this extended state-space model satisfies the following assumption, similar to Assumption 3.

Assumption 5 *The log-likelihood function $\bar{\ell}: \mathbb{R}^{dT} \to \mathbb{R}$ is four times continuously differentiable and, for δ defined as in Assumption 2, the associated likelihood $\overline{\mathcal{L}}: \mathbb{R}^{dT} \to \mathbb{R}$ satisfies:*

$$\forall \theta \in \mathbb{R}^d \quad \exists 0 < \eta < \delta \quad \exists \epsilon, D > 0 \quad \forall u_{1:T} \in \mathbb{R}^{dT} \quad \overline{\mathcal{L}}(\theta^{[T]} + u_{1:T}) \le De^{\epsilon\sum_{t=1}^{T}|u_t|^\eta}.$$

This assumption in conjunction with Assumptions 1 and 2 allows us to obtain the equivalent of Theorem 1 for $\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}\{h(\widetilde{\Theta}_{1:T} - \theta^{[T]})|\widetilde{Y}_{1:T} = y_{1:T}\}$ with $\overline{\kappa}(u_{1:T}) = \prod_{t=1}^{T} \kappa(u_t)$ and $\bar{\ell}^{(i)}(\theta^{[T]})$ in place of κ and $\ell^{(i)}(\theta)$. It is not stated here for the sake of brevity.

3.3 Approximation of the score vector and observed information matrix

It is now possible to adapt the results of Section 2.2 to obtain new estimates of the score and observed information matrix for state-space models.

Theorem 4 *For the artificial Bayesian model introduced in Section 3.2, satisfying Assumptions 1-2-5, for any $\theta \in \mathbb{R}^d$ there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \leq \eta$*

$$\left| \ell^{(1)}(\theta) - S_{\tau,T}(\theta) \right| \leq C\tau^2. \tag{15}$$

where the score estimator is given by

$$S_{\tau,T}(\theta) = \tau^{-2}\Sigma^{-1} \left\{ \sum_{t=1}^{T} \widetilde{\mathbb{E}}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} \right) - T\theta \right\}. \tag{16}$$

Remark 5 *In [13], the score estimate is obtained by considering a different stochastically perturbed state-space model where $\widetilde{\Theta}_0 \sim \tau^{-d}\kappa\left\{\tau^{-1}\left(\cdot - \theta_t\right)\right\},$*

$$\widetilde{\Theta}_t - \theta_t = \widetilde{\Theta}_{t-1} - \theta_{t-1} + \sigma V_t, \qquad V_t \overset{i.i.d.}{\sim} \kappa\left(\cdot\right) \tag{17}$$

and it is established that, for any compact K, there exist $\eta > 0$ and $C < \infty$ such that

$$\sup_{\theta \in K} |\ell^{(1)}(\theta) - \tau^{-2} \left\{ \widetilde{\mathbb{E}}_{\theta^{[T]},\tau,\sigma} \left(\widetilde{\Theta}_T \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \theta \right\} | \leq C \left(\tau + \frac{\sigma^2}{\tau^2} \right). \tag{18}$$

Note that the estimate in (18) requires solving a filtering problem whereas our estimate in (15) requires solving a smoothing problem.

As in Section 2.2 we complete this section with an approximation of the observed information matrix. To obtain simple approximations we assume that κ satisfies Assumption 4.

Theorem 6 *For the artificial Bayesian model introduced in Section 3.2, satisfying Assumptions 1-2-4-5, for any $\theta \in \mathbb{R}^d$, there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \leq \eta$*

$$\left| -\ell^{(2)}(\theta) - I_{\tau,T}(\theta) \right| \leq C\tau^2 \tag{19}$$

where the observed information matrix estimator is given by

$$I_{\tau,T}(\theta) = -\tau^{-4}\Sigma^{-1} \left\{ \sum_{s=1}^{T}\sum_{t=1}^{T} \widetilde{\mathbb{C}}ov_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 T\Sigma \right\} \Sigma^{-1}. \tag{20}$$

3.4 Sequential Monte Carlo estimates

The approximations of the score vector and observed information matrix given in Sections 3.3 require computing $\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}(\widetilde{\Theta}_t|\widetilde{Y}_{1:T} = y_{1:T})$ and $\widetilde{\mathbb{C}}ov_{\theta^{[T]},\tau}(\widetilde{\Theta}_s, \widetilde{\Theta}_t|\widetilde{Y}_{1:T} = y_{1:T})$ for $s, t \in (1, 2, ..., T)$. These can be approximated using sequential Monte Carlo methods applied to the modified state-space model described in (14) with $\theta_{1:T} = \theta^{[T]}$. This provides us with an approximation of $\widetilde{p}_{\widetilde{\Theta}_{1:T}, \widetilde{X}_{1:T}|\widetilde{Y}_{1:T}}(\tilde{\theta}_{1:T}, x_{1:T}|y_{1:T}; \theta^{[T]}, \tau)$, hence of its marginals $\widetilde{p}_{\widetilde{\Theta}_t|\widetilde{Y}_{1:T}}(\tilde{\theta}_t|y_{1:T}; \theta^{[T]}, \tau)$ and $\widetilde{p}_{\widetilde{\Theta}_s, \widetilde{\Theta}_t|\widetilde{Y}_{1:T}}(\tilde{\theta}_s, \tilde{\theta}_t|y_{1:T}; \theta^{[T]}, \tau)$. However, this approximation will be progressively impoverished as T increases because of the successive resampling steps. Eventually, $\widetilde{p}_{\widetilde{\Theta}_t|\widetilde{Y}_{1:T}}(\tilde{\theta}_t|y_{1:T}; \theta^{[T]}, \tau)$ will be approximated by a single unique particle for $T - t$ sufficiently large. To obtain lower variance estimators, it

is possible to use sequential Monte Carlo smoothing techniques. Standard approaches include the forward backward smoothing procedure [7] and the generalized two-filter smoothing recursion [3]. However these approaches are only applicable when we can evaluate $f(x' | x; \theta)$ pointwise and the primary motivation for this work is to address scenarios where this is not possible. In this case, we can only use the bootstrap filter [11] and can still obtain lower variance estimators at the cost of a small bias increase by using the fact that, when the state-space model enjoys forgetting properties, we have

$$\widetilde{\mathbb{E}}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \approx \widetilde{\mathbb{E}}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta) \wedge T} = y_{1:(t+\Delta) \wedge T} \right)$$

for a lag Δ large enough. This fixed-lag approximation was first proposed in [14] and has been studied in [16]. Similarly, w.l.o.g. consider that for $t \geq s$ we have

$$\widetilde{\mathbb{C}ov}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \approx \widetilde{\mathbb{C}ov}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta) \wedge T} = y_{1:(t+\Delta) \wedge T} \right)$$

and for $t - s > \Delta$

$$\widetilde{\mathbb{C}ov}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \approx 0.$$

This suggests to practically approximate using the bootstrap filter the following fixed-lag smoothing approximations of the score vector $S_{\tau, T}(\theta)$ and the observed information matrix $I_{\tau, T}(\theta)$:

$$S_{\tau, \Delta, T}(\theta) = \tau^{-2} \Sigma^{-1} \left\{ \sum_{t=1}^{T} \widetilde{\mathbb{E}}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta) \wedge T} = y_{1:(t+\Delta) \wedge T} \right) - T\theta \right\}, \tag{21}$$

$$I_{\tau, \Delta, T}(\theta) = -\tau^{-4} \Sigma^{-1} \left\{ \sum_{t=1}^{T} \widetilde{\mathbb{C}ov}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta) \wedge T} = y_{1:(t+\Delta) \wedge T} \right) \right. \tag{22}$$
$$\left. + 2 \sum_{s=1}^{T} \sum_{t=s+1}^{(s+\Delta) \wedge T} \widetilde{\mathbb{C}ov}_{\theta^{[T]}, \tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta) \wedge T} = y_{1:(t+\Delta) \wedge T} \right) - \tau^2 T \Sigma \right\} \Sigma^{-1}$$

with the convention that $\sum_{k=i}^{j} = 0$ if $i > j$.

3.5   Convergence results

We first quantify below the bias brought by the fixed-lag approximation. Our bounds rely on the following mixing assumption.

Assumption 6 (a) The set $\mathcal{S}(\theta) = \left[\widetilde{\theta} \in \mathbb{R}^d : \kappa \left\{ \left(\widetilde{\theta} - \theta \right) / \tau \right\} > 0 \right]$ is compact, so $d(\theta, \tau) := \sup_{\widetilde{\theta} \in \mathcal{S}(\theta)} \left| \widetilde{\theta} \right| < \infty$.

(b) $\lambda(dx)$ is a probability measure.

(c) $\underline{\alpha}(\theta) = \inf_{\widetilde{\theta}, x, x' \in \mathcal{S}(\theta) \times \mathcal{X} \times \mathcal{X}} f \left(x' | x; \widetilde{\theta} \right) > 0$, $\overline{\alpha}(\theta) = \sup_{\widetilde{\theta}, x, x' \in \mathcal{S}(\theta) \times \mathcal{X} \times \mathcal{X}} f \left(x' | x; \widetilde{\theta} \right) < \infty$ so $\rho(\theta) = 1 - \underline{\alpha}(\theta) / \overline{\alpha}(\theta) > 0$.

(d) for all $y \in \mathcal{Y}$, $\overline{g}(y; \theta) = \sup_{\widetilde{\theta}, x \in \mathcal{S}(\theta) \times \mathcal{X}} g \left(y | x; \widetilde{\theta} \right) < \infty$, $\underline{g}(y; \theta) = \int g \left(y | x; \widetilde{\theta} \right) \tau^{-d} \kappa \left\{ \tau^{-1} \left(\widetilde{\theta} - \theta \right) \right\} d\widetilde{\theta} \lambda(dx) > 0$, $\underline{g}_1(y; \theta) = \int g \left(y | x; \widetilde{\theta} \right) \tau^{-d} \kappa \left\{ \tau^{-1} \left(\widetilde{\theta} - \theta \right) \right\} d\widetilde{\theta} \, \nu(x; \theta) \lambda(dx) > 0$.

Weaker conditions could be used at the cost of substantially more complex proofs; see [6, chapter 4], [18].

Proposition 7 Suppose Assumption 6. Then for all integers $T \geq 1$, $0 \leq \Delta \leq T - 1$, we have

$$\tau^2 \left| \Sigma \left\{ S_{\tau, \Delta, T}(\theta) - S_{\tau, T}(\theta) \right\} \right| \leq 2 d(\theta, \tau) \rho(\theta)^{\Delta} (T - 1 - \Delta), \tag{23}$$

$$\tau^4 \left| \Sigma \left\{ I_{\tau, \Delta, T}(\theta) - I_{\tau, T}(\theta) \right\} \Sigma \right| \leq 2 d(\theta, \tau)^2 \rho(\theta)^{\Delta} (T - 1 - \Delta) \left[3 + 6\Delta + \frac{2\rho(\theta)}{1 - \rho(\theta)} \right]. \tag{24}$$

Let $S_{\tau, \Delta, T}^N(\theta)$ and $I_{\tau, \Delta, T}^N(\theta)$ be the bootstrap filter approximations of $S_{\tau, T}(\theta)$ and $I_{\tau, \Delta, T}(\theta)$ based on N particles. The following proposition relies partly on results in [16].

Proposition 8 *Suppose Assumption 6. Then for all integers $T \geq 1$, $0 \leq \Delta \leq T-1$, $N \geq 1$ and for any $p \geq 2$, there exist constants B and B_p, dependent only on p, such that*

$$\tau^2 \left| \mathbb{E} \left[\Sigma \left\{ S^N_{\tau,\Delta,T}(\theta) - S_{\tau,T}(\theta) \right\} \right] \right| \leq \frac{d(\theta,\tau)}{N} \sum_{t=1}^{T} C_t(\theta) \tag{25}$$

$$\tau^4 \left| \mathbb{E} \left[\Sigma \left\{ I^N_{\tau,\Delta,T}(\theta) - I_{\tau,\Delta,T}(\theta) \right\} \Sigma \right] \right| \leq \frac{d(\theta,\tau)^2}{N} \left[\sum_{t=1}^{T} \left\{ 3C_t(\theta) + D_t^2(\theta) \right\} + 2 \sum_{s=1}^{T} \sum_{t=s+1}^{(s+\Delta)\wedge T} \left\{ C_s(\theta) + 2C_t(\theta) + D_s^2(\theta) D_t^2(\theta) \right\} \right] \tag{26}$$

and

$$\tau^2 \mathbb{E}^{1/p} \left[\left| \Sigma \left\{ S^N_{\tau,\Delta,T}(\theta) - S_{\tau,\Delta,T}(\theta) \right\} \right|^p \right] \leq \frac{d(\theta,\tau)}{\sqrt{N}} \sum_{t=1}^{T} D_t^p(\theta) \tag{27}$$

$$\tau^4 \mathbb{E}^{1/p} \left[\left| \Sigma \left\{ I^N_{\tau,\Delta,T}(\theta) - I_{\tau,\Delta,T}(\theta) \right\} \Sigma \right|^p \right] \leq \frac{d(\theta,\tau)^2}{\sqrt{N}} \left[\sum_{t=1}^{T} 3D_t^p(\theta) + 2 \sum_{s=1}^{T} \sum_{t=s+1}^{(s+\Delta)\wedge T} \left\{ D_s^p(\theta) + 2D_t^p(\theta) \right\} \right] \tag{28}$$

where the expectations are with respect to the law of the bootstrap filter and

$$C_t(\theta) = \frac{B}{\underline{\alpha}(\theta)^2 \left\{ 1 - \rho(\theta) \right\}^2} \sum_{k=2}^{(t+\Delta)\wedge T} \frac{\overline{g}(y_k;\theta)^2}{\underline{g}(y_k;\theta)^2} \rho(\theta)^{0 \vee (t-k)} + \frac{\overline{g}_1(y_1;\theta)^2}{\underline{g}_1(y_1;\theta)^2} \rho(\theta)^{t-1},$$

$$D_t^p(\theta) = \frac{B_p}{\underline{\alpha}(\theta) \left\{ 1 - \rho(\theta) \right\}} \sum_{k=2}^{(t+\Delta)\wedge T} \frac{\overline{g}(y_k;\theta)}{\underline{g}(y_k;\theta)} \rho(\theta)^{0 \vee (t-k)} + \frac{\overline{g}_1(y_1;\theta)}{\underline{g}_1(y_1;\theta)} \rho(\theta)^{t-1} + 1.$$

4 Discussion

Building upon [12, 13], we have proposed a derivative-free estimator of the observed information matrix for general statistical models which can be computed easily using Bayesian computational tools. In the specific context of state-space models, we have also obtained new derivative-free estimators of the score vector and observed information matrix. These estimators are obtained by solving smoothing problems for a modified state-space model that differs from the one proposed in [12, 13]. Under mixing assumptions on the original state-space model, it is possible to obtain quantitative bounds for the resulting sequential Monte Carlo estimators.

Extensive numerical experiments comparing in practical situations the various estimators discussed in the paper will be made available shortly.

Acknowledgments

A Proofs

In the proofs we will use the notation

$$C.u^{\otimes k} = C \times \sum_{1 \leq i_1, \ldots, i_k \leq d} u_{i_1} \ldots u_{i_k}$$

for any $C \in \mathbb{R}$ and we will denote by $B(x,r)$ the ball of radius $r > 0$ centered at $x \in \mathbb{R}^d$ for the L_1 norm. Our calculations use extensively the elementary fact that, for κ a symmetric prior, for odd k and for any $1 \leq i_1, \ldots, i_k \leq d$

$$\int u_{i_1} \ldots u_{i_k} \, \kappa(u) du = 0. \tag{29}$$

The proof of Theorem 1 uses techniques borrowed from the literature on Bayesian asymptotic theory, see for instance the first chapter of [9] and references therein. Indeed both estimators provided by Theorems 2 and 3 are based on the asymptotic moments of the posterior distribution as it concentrates. However in our context, the posterior concentrates because the prior concentrates whereas the likelihood function is fixed. In Bayesian asymptotic theory, the posterior concentrates because the likelihood function concentrates and the prior is fixed. The proof of the main Theorem 1 relies on the following Proposition.

Proposition 9 *Suppose Assumptions 1-2-3. For any $\theta \in \mathbb{R}^d$, and $h : \mathbb{R}^d \to \mathbb{R}^m$ satisfying (4) for some constants $\alpha \geq 0$ and $c > 0$ we have for any $\tau > 0$*

$$\int h(\tau u)\frac{\mathcal{L}(\theta + \tau u)}{\mathcal{L}(\theta)}\kappa(u)du = \int h(\tau u)\kappa(u)du + \tau \int h(\tau u)\ell^{(1)}(\theta).u\ \kappa(u)du$$

$$+ \tau^2 \int h(\tau u)\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du$$

$$+ \tau^3 \int h(\tau u)\left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3 + \frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2}) + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\}\kappa(u)du + \mathcal{O}(\tau^{4+\alpha}).$$

Proof. We start by dividing the integral into two parts

$$\int h(\tau u)\frac{\mathcal{L}(\theta + \tau u)}{\mathcal{L}(\theta)}\kappa(u)du = \int_{\tau|u|\leq\rho} h(\tau u)\frac{\mathcal{L}(\theta + \tau u)}{\mathcal{L}(\theta)}\kappa(u)du + \int_{\tau|u|>\rho} h(\tau u)\frac{\mathcal{L}(\theta + \tau u)}{\mathcal{L}(\theta)}\kappa(u)du$$

for ρ in $\mathbb{R}^+$. We now fix ρ.

The expansion in Proposition 9 stems from the first part, while the second part will end up in the $\mathcal{O}(\tau^{4+\alpha})$ remainder. We look at these two terms separately.

First let us rewrite the first part of the integral as

$$\int_{\tau|u|\leq\rho} h(\tau u)\frac{\mathcal{L}(\theta + \tau u)}{\mathcal{L}(\theta)}\kappa(u)du = \int_{\tau|u|\leq\rho} h(\tau u)\exp\left\{\ell(\theta + \tau u) - \ell(\theta)\right\}\kappa(u)du.$$

We then use multiple Taylor expansions for a fixed value of τu. First, we have

$$\ell(\theta + \tau u) = \ell(\theta) + \tau\ell^{(1)}(\theta).u + \frac{\tau^2}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{\tau^3}{3!}\ell^{(3)}(\theta).u^{\otimes 3} + R_3(\theta, \tau u)$$

where $R_3(\theta, \tau u)$ simply denotes the remainder. Then using

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{3!}\int_0^1 (1-t)^3 e^{xt}dt,$$

we obtain

$$e^{\ell(\theta+\tau u)-\ell(\theta)} = 1 + \tau\ell^{(1)}(\theta).u + \frac{\tau^2}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{\tau^3}{3!}\ell^{(3)}(\theta).u^{\otimes 3} + R_3(\theta, \tau u) \qquad (30)$$

$$+ \frac{1}{2}\left\{\tau\ell^{(1)}(\theta).u + \frac{\tau^2}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{\tau^3}{3!}\ell^{(3)}(\theta).u^{\otimes 3} + R_3(\theta, \tau u)\right\}^2$$

$$+ \frac{1}{3!}\left\{\tau\ell^{(1)}(\theta).u + \frac{\tau^2}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{\tau^3}{3!}\ell^{(3)}(\theta).u^{\otimes 3} + R_3(\theta, \tau u)\right\}^3$$

$$+ + \frac{1}{3!}\left\{\tau\ell^{(1)}(\theta).u + \frac{\tau^2}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{\tau^3}{3!}\ell^{(3)}(\theta).u^{\otimes 3} + R_3(\theta, \tau u)\right\}^4 \int_0^1 (1-t)^3 e^{\{\ell(\theta+\tau u)-\ell(\theta)\}t}dt.$$

We then integrate this expression multiplied by $h(\tau u)\kappa(u)$ over u on the set $\{u : \tau|u| \leq \rho\}$, and group the terms as follows

$$\int_{\tau|u|\leq\rho} h(\tau u)e^{\ell(\theta+\tau u)-\ell(\theta)}\kappa(u)du = \int_{\tau|u|\leq\rho} h(\tau u)\kappa(u)du + \tau \int_{\tau|u|\leq\rho} h(\tau u)\ell^{(1)}(\theta).u\ \kappa(u)du$$

$$+ \tau^2 \int_{\tau|u|\leq\rho} h(\tau u)\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du$$

$$+ \tau^3 \int_{\tau|u|\leq\rho} h(\tau u)\left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3 + \frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2}) + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\}\kappa(u)du + Q(\theta, \tau)$$

9

where $Q(\theta, \tau)$ is the sum of all the remaining terms that were in (30). We would like to prove that $Q(\theta, \tau) = \mathcal{O}(\tau^{4+\alpha})$. Let us look at the various terms in this remainder.

1. First, some terms of this sum are of the form:

$$K\tau^{j_1+j_2+j_3} \int_{\tau|u|\leq\rho} h(\tau u) \left(\ell^{(1)}(\theta).u\right)^{j_1} \left(\ell^{(2)}(\theta).u^{\otimes 2}\right)^{j_2} \left(\ell^{(3)}(\theta).u^{\otimes 3}\right)^{j_3} \kappa(u) du \qquad (31)$$

for some $K > 0$ and non-negative integers j_1, j_2, j_3 such that $j_1 + j_2 + j_3 \geq 4$. Using (4) and Assumption 1 guaranteeing that κ has finite moments of all orders, we bound the integral in (31) by $L\tau^\alpha$ for some constant L independent of τ and hence terms like in (31) are $\mathcal{O}(\tau^{j_1+j_2+j_3+\alpha}) = \mathcal{O}(\tau^{4+\alpha})$.

2. Then, some terms are similar to (31) but also include powers of $R_3(\theta, \tau u)$ under the integral. We use the Lagrange form of the remainder $R_3(\theta, \tau u)$:

$$R_3(\theta, \tau u) = \frac{\tau^4}{4!}\ell^{(4)}(\theta^\star).u^{\otimes 4}$$

for some $\theta^\star \in B(\theta, \tau u)$. Since the log-likelihood ℓ has a continuous fourth derivative (Assumption 3) there is a constant $m(\theta, \rho) > 0$ independent of τ such that for every $\theta^\star \in B(\theta, \rho)$ and every $u \in \mathbb{R}^d$:

$$\left|\ell^{(4)}(\theta^\star).u^{\otimes 4}\right| < \left|m(\theta, \rho).u^{\otimes 4}\right|. \qquad (32)$$

Hence, we have

$$|R_3(\theta, \tau u)| < \frac{\tau^4 \left|m(\theta, \rho).u^{\otimes 4}\right|}{4!}.$$

Therefore the terms that include powers of $R_3(\theta, \tau u)$ are also $\mathcal{O}(\tau^{4+\alpha})$.

3. Finally, some terms of $Q(\theta, \tau)$ (the ones that come from the last line of (30)) also include the following expression under the integral over u

$$\int_0^1 (1-t)^3 e^{\{\ell(\theta+\tau u)-\ell(\theta)\}t} dt.$$

Since the log-likelihood ℓ is continuous (Assumption 3) there exists a constant $C(\theta, \rho) > 0$ independent of τ such that

$$\sup_{\theta^\star \in B(\theta, \rho)} |\ell(\theta^\star) - \ell(\theta)| < C(\theta, \rho). \qquad (33)$$

Hence we obtain for any u such that $\tau|u| \leq \rho$

$$\int_0^1 (1-t)^3 e^{\{\ell(\theta+\tau u)-\ell(\theta)\}t} dt \leq \int_0^1 (1-t)^3 e^{C(\theta,\rho)t} dt \leq \widetilde{C}(\theta, \rho)$$

for some $\widetilde{C}(\theta, \rho) \in \mathbb{R}$ and this integral does not cause additional difficulties.

Hence all terms in $Q(\theta, \tau)$ are $\mathcal{O}(\tau^{4+\alpha})$. At this point we have

$$\int_{\tau|u|\leq\rho} h(\tau u) e^{\ell(\theta+\tau u)-\ell(\theta)} \kappa(u) du = \int_{\tau|u|\leq\rho} h(\tau u) \kappa(u) du + \tau \int_{\tau|u|\leq\rho} h(\tau u) \ell^{(1)}(\theta).u \; \kappa(u) du$$

$$+ \tau^2 \int_{\tau|u|\leq\rho} h(\tau u) \left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\} \kappa(u) du$$

$$+ \tau^3 \int_{\tau|u|\leq\rho} h(\tau u) \left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3 + \frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2}) + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\} \kappa(u) du + \mathcal{O}(\tau^{4+\alpha}).$$

However we would like the integrals on the right-hand side of the equality sign to be over the whole space (as in the statement of Proposition 9) instead of being restricted to $\{u : \tau|u| \leq \rho\}$. Hence we want to add the following terms on the right-hand side

$$\int_{\tau|u|>\rho} h(\tau u) f^{[k]}(\theta).u^{\otimes k} \kappa(u) du \quad \text{for} \quad k = 0, 1, 2, 3$$

10

where $f^{[0]}(\theta).u^{\otimes 0} = 1$, $f^{[1]}(\theta).u^{\otimes 1} = \ell^{(1)}(\theta).u$, $f^{[2]}(\theta).u^{\otimes 2} = \frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2$ etc. We do this by proving that these integrals are $\mathcal{O}(\tau^{4+\alpha})$.

For τ small enough such that $\tau < \rho/M$ with M as in Assumption 2, we can write

$$\left| \int_{\tau|u|>\rho} h(\tau u) f^{[k]}(\theta).u^{\otimes k} \kappa(u) du \right| \le \int_{\tau|u|>\rho} c\tau^\alpha |u|^\alpha \left| f^{[k]}(\theta).u^{\otimes k} \right| e^{-\gamma|u|^\delta} du$$

and we can bound $\left| f^{[k]}(\theta).u^{\otimes k} \right|$ by $\left| f^{[k]}(\theta) \right| |u|^k$ where

$$\left| f^{[k]}(\theta) \right| := \sup_{|u|=1} f^{[k]}(\theta).u^{\otimes k} < \infty$$

to obtain

$$\left| \int_{\tau|u|>\rho} h(\tau u) f^{[k]}(\theta).u^{\otimes k} \kappa(u) du \right| \le c \left| f^{[k]}(\theta) \right| \tau^\alpha \int_{\tau|u|>\rho} |u|^{k+\alpha} e^{-\gamma|u|^\delta} du.$$

We will conclude by proving for any $m \in \mathbb{R}, \mu \in \mathbb{R}$

$$\int_{\tau|u|>\rho} |u|^m e^{-\gamma|u|^\delta} du = \mathcal{O}(\tau^\mu). \tag{34}$$

Let us prove (34). We have

$$|u|_2 \le |u| \le d^{1/2}|u|_2$$

where $|u|_2 = \left(\sum_{i=1}^d u_i^2 \right)^{1/2}$ is the Euclidean norm so

$$\int_{\tau|u|>\rho} |u|^m e^{-\gamma|u|^\delta} du \le d^{m/2} \int_{\tau|u|_2>\rho} |u|_2^m e^{-\gamma|u|_2^\delta} du.$$

First we change to spherical coordinates

$$\int_{\tau|u|_2>\rho} |u|_2^m e^{-\gamma|u|_2^\delta} du = \left(\int_{\tau r>\rho} r^{m+d-1} e^{-\gamma r^\delta} dr \right) \times S_{d-1}(1)$$

where r represents the radius, and $S_{d-1}(1)$ is the surface of the d-dimensional unit ball of radius 1 associated to the Euclidean norm. We now handle a simpler integral on the one-dimensional variable r:

$$\int_{\tau r>\rho} r^{m+d-1} e^{-\gamma r^\delta} dr$$

$$(\text{with } s = r^\delta) = \frac{1}{\delta} \int_{s>\left(\frac{\rho}{\tau}\right)^\delta} s^{\frac{m+d-\delta}{\delta}} e^{-\gamma s} ds$$

$$(\text{with } t = s\tau^\delta) = \frac{\tau^{-(m+d)}}{\delta} \int_{t>\rho^\delta} t^{\frac{m+d-\delta}{\delta}} e^{-\gamma t/\tau^\delta} dt$$

$$(\text{multiplying and dividing by } e^{-\gamma\rho^\delta/\tau^\delta}) = \frac{\tau^{-(m+d)}}{\delta} e^{-\gamma\rho^\delta/\tau^\delta} \int_{t>0} \left(t+\rho^\delta\right)^{\frac{m+d-\delta}{\delta}} e^{-\gamma t/\tau^\delta} dt$$

$$(\text{as long as } \tau \le 1 \text{ so that } \forall t>0 \; -\gamma t/\tau^\delta \le -\gamma t \;) \le \frac{\tau^{-(m+d)}}{\delta} e^{-\gamma\rho^\delta/\tau^\delta} \int_{t>0} \left(t+\rho^\delta\right)^{\frac{m+d-\delta}{\delta}} e^{-\gamma t} dt$$

$$(\text{for all } \mu \in \mathbb{R}) = \mathcal{O}(\tau^\mu).$$

This allows to conclude $\int_{\tau|u|>\rho} h(\tau u) f^{[k]}(\theta).u^{\otimes k} \kappa(u) du = \mathcal{O}(\tau^{4+\alpha})$ for any $k = 0, 1, 2, 3$ and we finally obtain the desired expansion

$$\int_{\tau|u|\le\rho} h(\tau u) e^{\ell(\theta+\tau u)-\ell(\theta)} \kappa(u) du = \int h(\tau u)\kappa(u) du + \tau \int h(\tau u)\ell^{(1)}(\theta).u \; \kappa(u) du$$

$$+ \tau^2 \int h(\tau u) \left\{ \frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2 \right\} \kappa(u) du$$

$$+ \tau^3 \int h(\tau u) \left\{ \frac{1}{3!}(\ell^{(1)}(\theta).u)^3 + \frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2}) + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3} \right\} \kappa(u) du + \mathcal{O}(\tau^{4+\alpha}).$$

The second part of the integral of interest

$$\int_{\tau|u|>\rho} h(\tau u)\frac{\mathcal{L}(\theta+\tau u)}{\mathcal{L}(\theta)}\kappa(u)du$$

ends up inside an $\mathcal{O}(\tau^{4+\alpha})$ term through the following reasoning (note that we do not use Taylor expansions for this part, since it is the part where τu is large). Using Assumptions 2 and 3 we have (with M as in Assumption 2):

$$\left|\int_{\tau|u|>\rho} h(\tau u)\frac{\mathcal{L}(\theta+\tau u)}{\mathcal{L}(\theta)}\kappa(u)du\right| \le \frac{c\tau^\alpha}{\mathcal{L}(\theta)}\int_{\tau|u|>\rho}|u|^\alpha\mathcal{L}(\theta+\tau u)\kappa(u)du \tag{35}$$

$$\text{(provided that } \tau < \frac{\rho}{M} \text{ and with } v = \tau u) \le \frac{cD}{\tau\mathcal{L}(\theta)}\int_{|v|>\rho}|v|^\alpha e^{\epsilon|v|^\eta-\gamma|v|^\delta\tau^{-\delta}}dv$$

$$\text{(switching to Euclidean norm)} \le \frac{cDd^{\alpha/2}}{\tau\mathcal{L}(\theta)}\int_{|v|_2>\rho}|v|^\alpha e^{\epsilon d^{\eta/2}|v|_2^\eta-\gamma|v|_2^\delta\tau^{-\delta}}dv$$

$$\text{(changing to spherical coordinates)} = \frac{cDd^{\alpha/2}S_{d-1}(1)}{\tau\mathcal{L}(\theta)}\int_{r>\rho}r^{\alpha+d-1}e^{\epsilon d^{\eta/2}r^\eta-\gamma r^\delta\tau^{-\delta}}dr.$$

Then we take advantage of the assumption $\eta < \delta$ to bound $\epsilon d^{\eta/2}r^\eta - \gamma r^\delta\tau^{-\delta}$ by $-ar^\eta\tau^{-\delta}$ for some $a > 0$, again for τ small enough. Indeed consider the expression $\epsilon d^{\eta/2}r^\eta - \gamma r^\delta\tau^{-\delta}$ for $\epsilon, \eta, \gamma, \delta > 0$ with $\eta < \delta$, on the set $\{r > \rho\}$. Then

$$\epsilon d^{\eta/2}r^\eta - \gamma r^\delta\tau^{-\delta} = r^\eta\left(\epsilon d^{\eta/2} - \frac{\gamma}{\tau^\delta}r^{\delta-\eta}\right)$$

$$\le r^\eta\left(\frac{\epsilon d^{\eta/2}\tau^\delta - \gamma\rho^{\delta-\eta}}{\tau^\delta}\right)$$

where we bounded $r^{\delta-\eta}$ by $\rho^{\delta-\eta}$ using $\delta-\eta > 0$. Then we take τ small enough so that $\epsilon d^{\eta/2}\tau^\delta - \gamma\rho^{\delta-\eta} < -a < 0$ for some $a > 0$. With such an a, we have $\epsilon d^{\eta/2}r^\eta - \gamma r^\delta\tau^{-\delta} \le -ar^\eta\tau^{-\delta}$.

We end up with the following integral

$$\int_{r>\rho}r^{\alpha+d-1}e^{-ar^\eta\tau^{-\delta}}dr.$$

We then use the same reasoning as in the end of the previous section (see (34)) to conclude that this integral is $\mathcal{O}(\tau^\mu)$ for any power μ. ∎

Proof. of Theorem 1. We have the identity

$$\widetilde{\mathbb{E}}_{\theta,\tau}\left\{h\left(\widetilde{\Theta}-\theta\right)\middle|\widetilde{Y}=y\right\} = \frac{\int h\left(\widetilde{\theta}-\theta\right)p_Y\left(y;\widetilde{\theta}\right)\tau^{-d}\kappa\left\{\left(\widetilde{\theta}-\theta\right)/\tau\right\}d\widetilde{\theta}}{\int p_Y\left(y;\widetilde{\theta}\right)\tau^{-d}\kappa\left\{\left(\widetilde{\theta}-\theta\right)/\tau\right\}d\widetilde{\theta}}$$

$$= \frac{\int h\left(\tau u\right)\exp\left\{\ell\left(\theta+\tau u\right)-\ell(\theta)\right\}\kappa\left(u\right)du}{\int\exp\left\{\ell\left(\theta+\tau u\right)-\ell(\theta)\right\}\kappa\left(u\right)du} \tag{36}$$

where we have used Bayes formula in the first line and a substitution $u = \tau^{-1}(\widetilde{\theta}-\theta)$ in the second line. The numerator corresponds to the expansion obtained in Proposition 9, while the denominator is a particular case of the numerator when h is the constant function $h : u \mapsto 1$. In this case, Proposition 9 yields with $\alpha = 0$:

$$\int\exp\left\{\ell\left(\theta+\tau u\right)-\ell(\theta)\right\}\kappa(u)du = 1 + 0 + \tau^2\int\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du + 0 + \mathcal{O}(\tau^4)$$

where the zeros come from using (29). Hence

$$\left[\int\exp\left\{\ell\left(\theta+\tau u\right)-\ell(\theta)\right\}\kappa(u)du\right]^{-1} = 1 - \tau^2\int h(\tau u)\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2} + \frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du + \mathcal{O}(\tau^4)$$

and finally

$$\widetilde{\mathbb{E}}_{\theta,\tau}\left\{h\left(\widetilde{\Theta}-\theta\right)\Big|\widetilde{Y}=y\right\}=\left[\int h(\tau u)\kappa(u)du+\tau\int h(\tau u)\ell^{(1)}(\theta).u\ \kappa(u)du\right.$$

$$+\tau^2\int h(\tau u)\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2}+\frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du$$

$$+\tau^3\int h(\tau u)\left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3+\frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2})+\frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\}\kappa(u)du+O(\tau^{4+\alpha})\right]$$

$$\times\left[1-\tau^2\int h(\tau u)\left\{\frac{1}{2}\ell^{(2)}(\theta).u^{\otimes 2}+\frac{1}{2}(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du+\mathcal{O}(\tau^4)\right].$$

Putting the terms of order $\mathcal{O}(\tau^{4+\alpha})$ together we obtain Theorem 1. ∎

Proof. of Theorem 2. This is a direct consequence from Theorem 1. For $h(u)=u$, we have $\alpha=1$, and using (29)

$$\left|\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)-\tau^2\int u\ \ell^{(1)}(\theta).u\ \kappa(u)du\right.$$

$$-\tau^4\int u\ \left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3+\frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2})+\frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\}\kappa(u)du$$

$$\left.+\frac{\tau^4}{2}\left[\int\left\{\ell^{(2)}(\theta).u^{\otimes 2}+(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du\right]\left\{\int u\ \ell^{(1)}(\theta).u\ \kappa(u)du\right\}\right|\le C'\tau^5.$$

Under assumptions 1 and 2, the integrals appearing in the τ^4 terms are upper bounded so there exist η and $C''<\infty$ such that for $0<\tau\le\eta$,

$$\left|\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)|-\tau^2\int u\ell^{(1)}(\theta).u\ \kappa(u)du\right|\le C''\tau^4.$$

We can now conclude by noticing that $\int uu^{\mathrm{T}}\kappa\left(u\right)du=\Sigma$ where Σ is defined in Assumption 1. ∎

The proof of Theorem 3 relies on the following Proposition.

Proposition 10 *Suppose Assumptions 1-2-3-4. For any $\theta\in\mathbb{R}^d$, there exist $\eta>0$ and $C<\infty$ such that for all $0<\tau\le\eta$*

$$\left|\widetilde{\mathbb{E}}_{\theta,\tau}\left\{\left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^T\Big|\widetilde{Y}=y\right\}-\tau^2\Sigma-\tau^4 A\circ\left(\ell^{(2)}(\theta)+\ell^{(1)}(\theta)\ell^{(1)}(\theta)^T\right)\right|\le C\tau^6 \qquad (37)$$

where $\circ$ is the Hadamard product (i.e. element-wise product) and

$$A=\begin{pmatrix}\sigma_1^4 & \sigma_1^2\sigma_2^2 & \cdots & \sigma_1^2\sigma_d^2\\ \sigma_2^2\sigma_1^2 & \sigma_2^4 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \sigma_{d-1}^2\sigma_d^2\\ \sigma_d^2\sigma_1^2 & \cdots & \sigma_d^2\sigma_{d-1}^2 & \sigma_d^4\end{pmatrix}. \qquad (38)$$

Proof. The result is established by using Theorem 1 for $h\left(u\right)=uu^{\mathrm{T}}$ and $\alpha=2$. Given the matrix norm we use, Theorem 1 still holds and yields

$$\left|\widetilde{\mathbb{E}}_{\theta,\tau}\left\{\left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^{\mathrm{T}}\Big|\widetilde{Y}=y\right\}-\tau^2\int uu^{\mathrm{T}}\kappa(u)du-\tau^3\int uu^{\mathrm{T}}\ \ell^{(1)}(\theta).u\ \kappa(u)du\right.$$

$$-\frac{\tau^4}{2}\int\left(uu^{\mathrm{T}}-\int uu^{\mathrm{T}}\kappa(u)du\right)\left\{\ell^{(2)}(\theta).u^{\otimes 2}+(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du$$

$$-\tau^5\int uu^{\mathrm{T}}\left\{\frac{1}{3!}(\ell^{(1)}(\theta).u)^3+\frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2})+\frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3}\right\}\kappa(u)du$$

$$\left.+\frac{\tau^5}{2}\left[\int\left\{\ell^{(2)}(\theta).u^{\otimes 2}+(\ell^{(1)}(\theta).u)^2\right\}\kappa(u)du\right]\left\{\int uu^{\mathrm{T}}\ell^{(1)}(\theta).u\ \kappa(u)du\right\}\right|\le C\tau^6.$$

$$(39)$$

13

We have $\int uu^{\mathrm{T}}\kappa(u)du = \Sigma$ and elementwise using (29)

$$\int u_i u_j \ \ell^{(1)}(\theta).u \ \kappa(u)du = 0,$$

$$\int u_i u_j \ \left\{ \frac{1}{3!}(\ell^{(1)}(\theta).u)^3 + \frac{1}{2}(\ell^{(1)}(\theta).u)(\ell^{(2)}(\theta).u^{\otimes 2}) + \frac{1}{3!}\ell^{(3)}(\theta).u^{\otimes 3} \right\} \kappa(u)du = 0,$$

so

$$\left| \widetilde{\mathbb{E}}_{\theta,\tau}\left\{ \left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^{\mathrm{T}} \Big| \widetilde{Y}=y \right\} - \tau^2\Sigma - \frac{\tau^4}{2}\int \left(uu^{\mathrm{T}}-\Sigma\right)\left\{ \ell^{(2)}(\theta).u^{\otimes 2} + (\ell^{(1)}(\theta).u)^2 \right\}\kappa(u)du \right| \le C\tau^6.$$

Now we look elementwise at the integral on the left hand side of the above equation. The element i,j of the term is equal to

$$\frac{\tau^4}{2}\int \left(u_i u_j - \sigma_{i,j}\right)\left\{ \ell^{(2)}(\theta).u^{\otimes 2} + (\ell^{(1)}(\theta).u)^2 \right\}\kappa(u)du$$

$$= \frac{\tau^4}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\ell_{k,l}^{(1,2)}(\theta)\left(\Lambda_{i,j,k,l} - \sigma_{i,j}\sigma_{k,l}\right)$$

where we use the notation

$$\ell_{k,l}^{(1,2)}(\theta) := \frac{\partial^2\ell}{\partial\theta_k\partial\theta_l}(\theta) + \frac{\partial\ell}{\partial\theta_k}(\theta)\frac{\partial\ell}{\partial\theta_l}(\theta)$$

and $\Lambda_{i,j,k,l} = \int u_i u_j u_k u_l \ \kappa(u)du$. Because of Assumption 4, the element i,i is equal to

$$\frac{\tau^4}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\ell_{k,l}^{(1,2)}(\theta)\left(\Lambda_{i,i,k,l} - \sigma_i^2\sigma_{k,l}\right) = \frac{\tau^4}{2}\ell_{i,i}^{(1,2)}(\theta)\left(\Lambda_i - \sigma_i^4\right) = \tau^4\sigma_i^4\ell_{i,i}^{(1,2)}(\theta)$$

since $\Lambda_{i,i,k,k} = \sigma_i^2\sigma_k^2$ for $i \ne k$. The element i,j, for $i \ne j$ is equal to

$$\frac{\tau^4}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\ell_{k,l}^{(1,2)}(\theta)\left(\Lambda_{i,j,k,l} - \sigma_{i,j}\sigma_{k,l}\right) = \frac{\tau^4}{2}\sum_{k=1}^{d}\sum_{l=1}^{d}\ell_{k,l}^{(1,2)}(\theta)\Lambda_{i,j,k,l} = \tau^4\ell_{i,j}^{(1,2)}(\theta)\sigma_i^2\sigma_j^2$$

as $\Lambda_{i,i,k,l} = \sigma_i^2\sigma_{k,l}$ when i,j,k are distinct. The result of the proposition follows. ∎

Proof. of Theorem 3. Proposition 10 yields

$$\left| \left\{ \ell^{(2)}(\theta) + \ell^{(1)}(\theta)\ell^{(1)}(\theta)^{\mathrm{T}} \right\} - \tau^{-4}B \circ \left[\widetilde{\mathbb{E}}_{\theta,\tau}\left\{ \left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^{\mathrm{T}} \Big| \widetilde{Y}=y \right\} - \tau^2\Sigma \right] \right| \le C'\tau^2 \qquad (40)$$

where $B \in \mathbb{R}^{d\times d}$ is the matrix such that $B_{ij} = A_{ij}^{-1}$ with A given by (38). We also have

$$\left| \ell^{(1)}(\theta)\ell^{(1)}(\theta)^{\mathrm{T}} - \tau^{-4}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)^{\mathrm{T}}\Sigma^{-1} \right|$$

$$= \left| \left\{ \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right\}\ell^{(1)}(\theta)^{\mathrm{T}} \right.$$

$$\left. + \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)\left\{ \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right\}^{\mathrm{T}} \right|$$

$$\le \left| \left\{ \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right\}\ell^{(1)}(\theta)^{\mathrm{T}} \right|$$

$$+ \left| \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right)\left\{ \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right\}^{\mathrm{T}} \right|$$

$$\le \left| \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right|\left(\left| \ell^{(1)}(\theta) \right| + \tau^{-2}\left| \Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right| \right).$$

Using Theorem 2 yields

$$\left| \ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\Big|\widetilde{Y}=y\right) \right| \le C''\ \tau^2$$

14

for some constant C'', and using the triangle inequality

$$\left|\tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)\right| \leq \left|\ell^{(1)}(\theta)\right| + \left|\ell^{(1)}(\theta)-\tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)\right| \leq \left|\ell^{(1)}(\theta)\right| + C''\,\tau^2$$

which leads to

$$\left|\ell^{(1)}(\theta)\ell^{(1)}(\theta)^{\mathrm{T}} - \tau^{-4}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)^{\mathrm{T}}\Sigma^{-1}\right| \leq C'''\,\tau^2. \tag{41}$$

for some constant C''' when τ is small enough. Combining (40) and (41) we obtain

$$\left|\ell^{(2)}(\theta) - \tau^{-4}\left[B\circ\left[\widetilde{\mathbb{E}}_{\theta,\tau}\left\{\left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^{\mathrm{T}}\middle|\widetilde{Y}=y\right\} - \tau^2\Sigma\right]\right.\right.$$
$$\left.\left.- \Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)^{\mathrm{T}}\Sigma^{-1}\right]\right| \leq C\tau^2. \tag{42}$$

Now using Assumption 4,

$$B\circ\left[\widetilde{\mathbb{E}}_{\theta,\tau}\left\{\left(\widetilde{\Theta}-\theta\right)\left(\widetilde{\Theta}-\theta\right)^{\mathrm{T}}\middle|\widetilde{Y}=y\right\} - \tau^2\Sigma\right] - \Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}-\theta\middle|\widetilde{Y}=y\right)^{\mathrm{T}}\Sigma^{-1}$$
$$= \Sigma^{-1}\left\{\widetilde{\mathbb{C}ov}_{\theta,\tau}\left(\widetilde{\Theta}\middle|\widetilde{Y}=y\right) - \tau^2\Sigma\right\}\Sigma^{-1}.$$

Indeed for the i-th diagonal term we obtain

$$(\sigma_i^4)^{-1}\left\{\widetilde{\mathbb{E}}_{\theta,\tau}\left(\left(\widetilde{\Theta}_i-\theta_i\right)^2\middle|\widetilde{Y}=y\right) - \tau^2\sigma_i^2\right\} - (\sigma_i^4)^{-1}\left\{\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}_i-\theta_i\middle|\widetilde{Y}=y\right)\right\}^2$$
$$= (\sigma_i^4)^{-1}\left\{\widetilde{\mathbb{V}}_{\theta,\tau}\left(\widetilde{\Theta}_i\middle|\widetilde{Y}=y\right) - \tau^2\sigma_i^2\right\}$$

while the (i,j) off-diagonal terms become

$$(\sigma_i^2\sigma_j^2)^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left\{\left(\widetilde{\Theta}_i-\theta_i\right)\left(\widetilde{\Theta}_j-\theta_j\right)\middle|\widetilde{Y}=y\right\} - (\sigma_i^2\sigma_j^2)^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}_i-\theta_i\middle|\widetilde{Y}=y\right)\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta}_j-\theta_j\middle|\widetilde{Y}=y\right)$$
$$= (\sigma_i^2\sigma_j^2)^{-1}\widetilde{\mathbb{C}ov}_{\theta,\tau}\left(\widetilde{\Theta}_i,\widetilde{\Theta}_j\middle|\widetilde{Y}=y\right)$$

which concludes the proof. ∎

Proof. of Theorem 4. It is straightforward to check that Theorem 2 holds for the artificial Bayesian model of Section 3.2 under assumptions 1, 2 and 5. Assumption 5 is necessary to ensure that developments as in (35) can be performed on the extended model. Hence for any $\theta \in \mathbb{R}^d$ there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \leq \eta$

$$|\overline{\ell}^{(1)}\left(\theta^{[T]}\right) - \tau^{-2}\bar{\Sigma}_T^{-1}\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}\left(\widetilde{\Theta}_{1:T}-\theta^{[T]}\middle|\widetilde{Y}_{1:T}=y_{1:T}\right)| \leq C\,\tau^2.$$

To prove the theorem we have to relate the dT-dimensional gradient $\overline{\ell}^{(1)}\left(\theta^{[T]}\right)$ to the desired d-dimensional $\ell^{(1)}(\theta)$. We have $\ell(\theta) = \overline{\ell}(\theta^{[T]})$ and the chain rules yields

$$\ell^{(1)}(\theta) = \sum_{t=1}^{T}\overline{\ell}_t^{(1)}(\theta^{[T]})$$

where $\overline{\ell}_t(\theta_{1:T})$ denotes $\partial\overline{\ell}(\theta_{1:T})/\partial\theta_t$. We have

$$|\sum_{t=1}^{T}\overline{\ell}_t^{(1)}(\theta^{[T]}) - \tau^{-2}\sum_{t=1}^{T}\left\{\bar{\Sigma}_T^{-1}\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}\left(\widetilde{\Theta}_{1:T}-\theta^{[T]}\middle|\widetilde{Y}_{1:T}=y_{1:T}\right)\right\}_t|$$
$$\leq \sum_{t=1}^{T}\left|\overline{\ell}_t^{(1)}(\theta^{[T]}) - \tau^{-2}\left\{\bar{\Sigma}_T^{-1}\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}\left(\widetilde{\Theta}_{1:T}-\theta^{[T]}\middle|\widetilde{Y}_{1:T}=y_{1:T}\right)\right\}_t\right|$$
$$= |\overline{\ell}^{(1)}(\theta^{[T]}) - \tau^{-2}\bar{\Sigma}_T^{-1}\widetilde{\mathbb{E}}_{\theta^{[T]},\tau}\left(\widetilde{\Theta}_{1:T}-\theta^{[T]}\middle|\widetilde{Y}_{1:T}=y_{1:T}\right)| \leq C\,\tau^2$$

15

where $\{s\}_t$ denotes the entries $\{d(t-1)+1,...,dt\}$ of a vector $s \in \mathbb{R}^{dT}$. Now because of the block structure of $\bar{\Sigma}_T^{-1}$, we have

$$\sum_{t=1}^{T} \left\{ \bar{\Sigma}_T^{-1} \widetilde{\mathbb{E}}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} - \theta^{[T]} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) \right\}_t = \Sigma^{-1} \widetilde{\mathbb{E}}_{\theta^{[T]},\tau} \left\{ \sum_{t=1}^{T} \left(\widetilde{\Theta}_t - \theta \right) \Big| \widetilde{Y}_{1:T} = y_{1:T} \right\}$$

and the result follows. ∎

Proof. of Theorem 6. Using the assumptions, Theorem 3 holds for the extended model. Hence for any $\theta \in \mathbb{R}^d$ there exist $\eta > 0$ and $C < \infty$ such that for all $0 < \tau \le \eta$

$$|\overline{\ell}^{(2)}(\theta^{[T]}) - \tau^{-4}\bar{\Sigma}_T^{-1} \left\{ \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 \bar{\Sigma}_T \right\} \bar{\Sigma}_T^{-1}| \le C\,\tau^2.$$

To prove the theorem we have to relate the $dT \times dT$-dimensional Hessian $\overline{\ell}^{(2)}(\theta^{[T]})$ to the desired $d \times d$-dimensional Hessian $\ell^{(2)}(\theta)$. The chain rule yields

$$\ell^{(2)}(\theta) = \sum_{s=1}^{T} \sum_{t=1}^{T} \overline{\ell}_{s,t}^{(2)}(\theta^{[T]})$$

where $\overline{\ell}_{s,t}^{(2)}(\theta_{1:T})$ denotes $\partial^2 \overline{\ell}(\theta_{1:T})/\partial \theta_s \partial \theta_t$. Hence we get for some $\eta > 0$, $C < \infty$ and any $0 < \tau \le \eta$

$$|\sum_{t=1}^{T} \sum_{s=1}^{T} \overline{\ell}_{s,t}^{(2)}(\theta^{[T]}) - \tau^{-4} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\bar{\Sigma}_T^{-1} \left\{ \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 \bar{\Sigma}_T \right\} \bar{\Sigma}_T^{-1} \right]_{t,s} |$$

$$\le \sum_{t=1}^{T} \sum_{s=1}^{T} \left| \overline{\ell}_{s,t}^{(2)}(\theta^{[T]}) - \tau^{-4} \left[\bar{\Sigma}_T^{-1} \left\{ \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 \bar{\Sigma}_T \right\} \bar{\Sigma}_T^{-1} \right]_{t,s} \right|$$

$$= |\overline{\ell}^{(2)}(\theta^{[T]}) - \tau^{-4} \bar{\Sigma}_T^{-1} \left\{ \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 \bar{\Sigma}_T \right\} \bar{\Sigma}_T^{-1}| \le C\,\tau^2$$

where $[M]_{t,s}$ denotes the (t,s) d-dimensional block of a matrix $M \in \mathbb{R}^{dT \times dT}$, and by using the L^1-norm on matrices. Under our assumptions $\bar{\Sigma}_T$ is diagonal and each of its diagonal blocks is equal to Σ, therefore we obtain

$$\sum_{t=1}^{T} \sum_{s=1}^{T} \left[\bar{\Sigma}_T^{-1} \left\{ \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 \bar{\Sigma}_T \right\} \bar{\Sigma}_T^{-1} \right]_{t,s}$$

$$= \Sigma^{-1} \left[\sum_{t=1}^{T} \sum_{s=1}^{T} \left[\widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_{1:T} \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) \right]_{t,s} - \tau^2 T\Sigma \right] \Sigma^{-1}$$

$$= \Sigma^{-1} \left\{ \sum_{s=1}^{T} \sum_{t=1}^{T} \widetilde{\mathbb{C}ov}_{\theta^{[T]},\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \Big| \widetilde{Y}_{1:T} = y_{1:T} \right) - \tau^2 T\Sigma \right\} \Sigma^{-1}$$

and the results follows. ∎

Proof. of Proposition 7. From Assumption (6), the transition kernel of the latent Markov state $\left(\widetilde{\Theta}_t, X_t \right)$ given $\left(\widetilde{\Theta}_{t-1}, X_{t-1} \right) = \left(\widetilde{\theta}_t, x_{t-1} \right)$ satisfies

$$\underline{\alpha}(\theta)\,\mu\left(d\widetilde{\theta}_t, dx_t \right) \le \tau^{-d} \kappa\left(\frac{\widetilde{\theta}_t - \theta}{\tau} \right) d\widetilde{\theta}_t\, f\left(x_t | x_{t-1}; \widetilde{\theta}_t \right) \lambda(dx_t) \le \overline{\alpha}(\theta)\,\mu\left(d\widetilde{\theta}_t, dx_t \right)$$

where

$$\mu\left(d\widetilde{\theta}_t, dx_t \right) := \tau^{-d} \kappa\left(\frac{\widetilde{\theta}_t - \theta}{\tau} \right) d\widetilde{\theta}_t\, \lambda(dx_t).$$

This ensure that the forward and backward smoothing kernels $\left\{ \widetilde{\mathbb{P}}_{\theta^{[T]},\tau} \left(d\widetilde{\theta}_{t+1}, dx_{t+1} \Big| \widetilde{\Theta}_t = \widetilde{\theta}_t, \widetilde{X}_t = x_t, \widetilde{Y}_{1:T} = y_{1:T} \right) \right\}_{t=1}^{T-1}$ and $\left\{ \widetilde{\mathbb{P}}_{\theta^{[T]},\tau} \left(d\widetilde{\theta}_t, dx_t \Big| \widetilde{\Theta}_{t+1} = \widetilde{\theta}_{t+1}, \widetilde{X}_{t+1} = x_{t+1}, \widetilde{Y}_{1:T} = y_{1:T} \right) \right\}_{t=1}^{T-1}$ are uniformly ergodic with mixing constant

$\rho(\theta)$; see e.g. [6, chapter 4], [16, Theorem 3.1]. For the score, we thus easily obtain that

$$\tau^2 \left| \Sigma \left\{ S_{\tau,\Delta,T}(\theta) - S_{\tau,T}(\theta) \right\} \right|$$

$$\leq \sum_{t=1}^{T-1-\Delta} \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta)\wedge T} = y_{1:(t+\Delta)\wedge T} \right) \right|$$

$$\leq 2d(\theta,\tau) \, \rho(\theta)^{\Delta} \, (T-1-\Delta)$$

where the last inequality follows from the ergodicity of the backward smoothing kernels. This yields (23). We now write $t(\Delta)$ for $(t+\Delta)\wedge T$. For the covariance, we have

$$\tau^4 \left| \Sigma \left\{ I_{\tau,\Delta,T}(\theta) - I_{\tau,T}(\theta) \right\} \Sigma \right|$$

$$\leq 2 \left| \sum_{s=1}^{T} \sum_{t=1+(s+\Delta)\wedge T}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right|$$

$$+ \left| \left\{ \sum_{t=1}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right. \right.$$

$$\left. \left. + 2 \sum_{s=1}^{T} \sum_{t=s+1}^{s(\Delta)} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right\} \right|.$$

For the first term we have

$$\left| \sum_{s=1}^{T} \sum_{t=1+s(\Delta)}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right| = \left| \sum_{s=1}^{T-\Delta-1} \sum_{t=1+(s+\Delta)\wedge T}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right|$$

$$\leq \sum_{s=1}^{T-\Delta-1} \sum_{t=1+s(\Delta)}^{T} \left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right|$$

where

$$\left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right|$$

$$= \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left\{ \widetilde{\Theta}_s \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t^{\mathrm{T}} \middle| \widetilde{X}_s, \widetilde{\Theta}_s, \widetilde{Y}_{1:T} = y_{1:T} \right) \middle| y_{1:T} \right\} - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \widetilde{\mathbb{E}}_{\theta[T],\tau}^{\mathrm{T}} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right|$$

$$= \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left[\widetilde{\Theta}_s \left\{ \mathbb{E} \left(\widetilde{\Theta}_t^{\mathrm{T}} \middle| \widetilde{X}_s, \widetilde{\Theta}_s, \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t^{\mathrm{T}} \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right\} \middle| \widetilde{Y}_{1:T} = y_{1:T} \right] \right|$$

$$\leq \widetilde{\mathbb{E}}_{\theta[T],\tau} \left[\left| \widetilde{\Theta}_s \left\{ \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{X}_s, \widetilde{\Theta}_s, \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right| \right\} \right| \middle| \widetilde{Y}_{1:T} = y_{1:T} \right]$$

$$\leq \widetilde{\mathbb{E}}_{\theta[T],\tau} \left[\left| \widetilde{\Theta}_s \right| \middle| \widetilde{Y}_{1:T} = y_{1:T} \right] 2d(\theta,\tau) \, \rho^{t-s} \leq 2d(\theta,\tau)^2 \, \rho(\theta)^{t-s}$$

where the last line of inequalities follows from the ergodicity of the forward smoothing kernels. Hence we have

$$2 \sum_{s=1}^{T} \sum_{t=1+s(\Delta)}^{T} \left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right| \leq 4d(\theta,\tau)^2 \sum_{s=1}^{T-\Delta-1} \sum_{t=1+s(\Delta)}^{T} \rho(\theta)^{t-s}$$

$$\leq 4d(\theta,\tau)^2 \, \frac{\rho(\theta)^{1+\Delta}}{1-\rho(\theta)} \, (T-\Delta-1). \tag{43}$$

We are now interested in upper bounding

$$
\left| \sum_{t=1}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta)\wedge T} = y_{1:(t+\Delta)\wedge T} \right) \right.
$$

$$
\left. + 2 \sum_{s=1}^{T} \sum_{t=s+1}^{(s+\Delta)\wedge T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta)\wedge T} = y_{1:(t+\Delta)\wedge T} \right) \right|
$$

$$
\leq \sum_{t=1}^{T} \left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta)\wedge T} = y_{1:(t+\Delta)\wedge T} \right) \right|
$$

$$
+ 2 \sum_{s=1}^{T} \sum_{t=s+1}^{(s+\Delta)\wedge T} \left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:(t+\Delta)\wedge T} = y_{1:(t+\Delta)\wedge T} \right) \right|.
$$

We have for $s \leq t$

$$
\left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
\leq \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \widetilde{\Theta}_t^{\mathrm{T}} \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \widetilde{\Theta}_t^{\mathrm{T}} \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
+ \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) \right| \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
+ \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_s \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right| \left| \widetilde{\mathbb{E}}_{\theta[T],\tau} \left(\widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
\leq 6 d \left(\theta, \tau \right)^2 \rho \left(\theta \right)^{\Delta}
$$

So we have

$$
\left| \sum_{t=1}^{T} \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
+ 2 \sum_{s=1}^{T} \sum_{t=s+1}^{s(\Delta)} \left| \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:T} = y_{1:T} \right) - \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right|
$$

$$
\leq 6 d \left(\theta, \tau \right)^2 \rho \left(\theta \right)^{\Delta} + 12 d \left(\theta, \tau \right)^2 \rho \left(\theta \right)^{\Delta} \Delta. \tag{44}
$$

The bound (24) follows by adding (43) to (44). ∎

Proof. of Proposition 8. The bounds on the score vector given in (27) and (25) follow directly from Proposition A1 in [16]. The bounds on the observed information matrix estimator are obtained as follows. We use $\widehat{\mathbb{E}}_{\theta[T],\tau}$ and $\widehat{\mathbb{C}ov}_{\theta[T],\tau}$ for the bootstrap filter approximations of expectation and covariance. Hence we have

$$
\tau^4 \left| \mathbb{E} \left[\Sigma \left\{ I_{\tau,\Delta,T}^{N} \left(\theta \right) - I_{\tau,\Delta,T} \left(\theta \right) \right\} \Sigma \right] \right|
$$

$$
= \sum_{t=1}^{T} \left| \mathbb{E} \left\{ \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_t, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right\} \right|
$$

$$
+ 2 \sum_{s=1}^{T} \sum_{t=s+1}^{s(\Delta)} \left| \mathbb{E} \left\{ \widetilde{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau} \left(\widetilde{\Theta}_s, \widetilde{\Theta}_t \middle| \widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)} \right) \right\} \right|
$$

where, using $ab - \widehat{a}\widehat{b} = a\left(b - \widehat{b}\right) + (a - \widehat{a})\left(\widehat{b} - b\right) + b\left(a - \widehat{a}\right)$, we have

$$\left|\mathbb{E}\left\{\widetilde{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_s, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_t, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$\leq \left|\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$+ \left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$+ \left|\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|$$

$$+ \left|\mathbb{E}\left[\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right.\right.$$

$$\times \left.\left.\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right]\right|$$

$$\leq \left|\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$+ d\left(\theta,\tau\right)\left|\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$+ d\left(\theta,\tau\right)\left|\mathbb{E}\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}\right|$$

$$+ \mathbb{E}\left[\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}^2\right]^{1/2}$$

$$\times \mathbb{E}\left[\left\{\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right\}^2\right]^{1/2}.$$

Hence we have using Proposition A1 from [16] that

$$\tau^4\left|\mathbb{E}\left[\Sigma\left\{I_{\tau,\Delta,T}^N\left(\theta\right) - I_{\tau,\Delta,T}\left(\theta\right)\right\}\Sigma\right]\right|$$

$$\leq d\left(\theta,\tau\right)^2\left[\sum_{t=1}^{T}\left\{3C_t\left(\theta\right) + D_t^2\left(\theta\right)\right\} + 2\sum_{s=1}^{T}\sum_{t=s+1}^{s(\Delta)}\left\{C_s\left(\theta\right) + 2C_t\left(\theta\right) + D_s^2\left(\theta\right)D_t^2\left(\theta\right)\right\}\right].$$

The bound (26) follows. Similarly we have by Minkowski's inequality

$$\tau^4\mathbb{E}^{1/p}\left[\left|\Sigma\left\{I_{\tau,\Delta,T}^N\left(\theta\right) - I_{\tau,\Delta,T}\left(\theta\right)\right\}\Sigma\right|^p\right]$$

$$= \sum_{t=1}^{T}\mathbb{E}\left[\left|\widetilde{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$+ 2\sum_{s=1}^{T}\sum_{t=s+1}^{s(\Delta)}\mathbb{E}\left[\left|\widetilde{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_s, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_s, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

where, using $ab - \widehat{a}\widehat{b} = a\left(b - \widehat{b}\right) + (a - \widehat{a})\widehat{b}$, we have

$$\mathbb{E}\left[\left|\widetilde{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_s, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{C}ov}_{\theta[T],\tau}\left(\widetilde{\Theta}_s, \widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$\leq \mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$+ \mathbb{E}\left[\left|\widetilde{\mathbb{E}}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:(t+\Delta)\wedge T}\right)\right|^p\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$+ \mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\left|\widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:(t+\Delta)\wedge T}\right)\right|^p\right]^{1/p}$$

$$\leq \mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\widetilde{\Theta}_t^{\mathrm{T}}\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$+ d\left(\theta,\tau\right)\mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_t\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}$$

$$+ d\left(\theta,\tau\right)\mathbb{E}\left[\left|\widetilde{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right) - \widehat{\mathbb{E}}_{\theta[T],\tau}\left(\widetilde{\Theta}_s\Big|\widetilde{Y}_{1:t(\Delta)} = y_{1:t(\Delta)}\right)\right|^p\right]^{1/p}.$$

Hence the bound (28) follows Proposition A1 in [16] x

$$\tau^4 \mathbb{E}^{1/p}\left[\left|\Sigma\left\{I_{\tau,\Delta,T}^N(\theta) - I_{\tau,\Delta,T}(\theta)\right\}\Sigma\right|^p\right] \le d\left(\theta,\tau\right)^2\left[\sum_{t=1}^{T} 3D_t^p(\theta) + 2\sum_{s=1}^{T}\sum_{t=s+1}^{s(\Delta)}\left\{D_s^p(\theta) + 2D_t^p(\theta)\right\}\right].$$

∎

B Optimal convergence rates

The estimators

$$\widehat{\ell}_{N,\tau}^{(1)}(\theta) = \tau^{-2}\Sigma^{-1}\widehat{\mu}_{N,\tau}(\theta), \quad \widehat{\ell}_{N,\tau}^{(2)}(\theta) = \tau^{-4}\Sigma^{-1}\left\{\widehat{v}_{N,\tau}(\theta) - \tau^2\Sigma\right\}\Sigma^{-1},$$

where $\widehat{\mu}_{N,\tau}(\theta)$ (resp. $\widehat{v}_{N,\tau}(\theta)$) is an importance sampling estimate of $\widetilde{\mathbb{E}}_{\theta,\tau}(\widetilde{\Theta} - \theta|\widetilde{Y} = y)$ (resp. $\widetilde{\mathbb{C}ov}_{\theta,\tau}(\widetilde{\Theta}|\widetilde{Y} = y)$), verify

$$\mathbb{E}\{\widehat{\mu}_{N,\tau}(\theta)\} = \widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta} - \theta\middle|\widetilde{Y} = y\right) + \frac{a}{N}, \quad \mathbb{V}\{\widehat{\mu}_{N,\tau}(\theta)\} = \frac{b}{N}\tau^2,$$

$$\mathbb{E}\{\widehat{v}_{N,\tau}(\theta)\} = \widetilde{\mathbb{C}ov}_{\theta,\tau}\left(\widetilde{\Theta}\middle|\widetilde{Y} = y\right) + \frac{c}{N}, \quad \mathbb{V}\{\widehat{v}_{N,\tau}(\theta)\} = \frac{d}{N}\tau^4,$$

for some a, b, c, d independent of τ^2 and N. These conditions are for example verified asymptotically in N if $\widehat{\mu}_{N,\tau}(\theta)$ and $\widehat{v}_{N,\tau}(\theta)$ are importance sampling estimator using the artificial prior as importance distribution. Hence the mean squared error for $\ell^{(1)}(\theta)$ satisfies

$$\mathbb{E}\left\{\tau^{-2}\Sigma^{-1}\widehat{\mu}_{N,\tau}(\theta) - \ell^{(1)}(\theta)\right\}^2$$

$$= \tau^{-4}\Sigma^{-2}\mathbb{V}\{\widehat{\mu}_{N,\tau}(\theta)\} + \left\{\ell^{(1)}(\theta) - \tau^{-2}\Sigma^{-1}\widetilde{\mathbb{E}}_{\theta,\tau}\left(\widetilde{\Theta} - \theta\middle|\widetilde{Y} = y\right)\right\}^2$$

$$\le \frac{e}{\tau^2 N} + f\tau^4$$

for some $e, f > 0$. This upper bound is minimised for τ of order $N^{-1/6}$ and is then of order $N^{-2/3}$. Similarly the mean squared error for $\ell^{(2)}(\theta)$ is

$$\mathbb{E}\left[\tau^{-4}\Sigma^{-2}\left\{\widehat{v}_{N,\tau}(\theta) - \tau^2\Sigma\right\} - \ell^{(2)}(\theta)\right]^2$$

$$= \tau^{-8}\Sigma^{-4}\mathbb{V}(\widehat{v}_{N,\tau}(\theta)) + \left[\ell^{(2)}(\theta) - \tau^{-4}\Sigma^{-1}\left\{\widetilde{\mathbb{C}ov}_{\theta,\tau}\left(\widetilde{\Theta}\middle|\widetilde{Y} = y\right) - \tau^2\Sigma\right\}\Sigma^{-1}\right]^2$$

$$\le \frac{g}{\tau^4 N} + h\tau^4$$

for some $g, h > 0$. The upper bound is minimised when τ is of order $N^{-1/8}$ and is then of order $N^{-1/2}$. Hence we find these estimators have the same optimal rate of convergence in terms of the mean squared error as the finite difference estimators.

References

[1] Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. Royal Stat. Soc.* B (with discussion), **72**, 269-342.

[2] Asmussen, S. and Glynn, P. (2007). *Stochastic Simulation: Algorithms and Analysis.* New York: Springer-Verlag.

[3] Briers, M., Doucet, A. and Maskell, S. (2010) Smoothing algorithms for state-space models. *Ann. Instit. Statist. Math.*, 62, 61-89.

[4] Cappé, O., Moulines, E. and Rydén, T. (2005) *Inference in Hidden Markov Models.* New York: Springer-Verlag.

[5] Chopin, N., Jacob, P.E. and Papaspiliopoulos, O. (2012). SMC^2: an efficient algorithm for sequential analysis of state space models. *J. Royal Stat. Soc.* B, **74**.

[6] Del Moral, P. (2004) *Feynman-Kac Formulae*. New York: Springer.

[7] Doucet, A., Godsill, S. J. and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10, 197–208.

[8] Doucet, A., De Freitas, J.F.G. and Gordon, N.J. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

[9] Ghosh, J.K. and Ramamoorthi, R.V. (2003). *Bayesian Nonparametrics*. New York: Springer-Verlag.

[10] Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. Royal Stat. Soc.* B (with discussion), **73(2)**, 123-214.

[11] Gordon, N.J., Salmond, D. and Smith, A.F.M. (1993). Novel approach to nonlinear non-Gaussian Bayesian state estimation. *IEE Proceedings* F, **40**, 107-113.

[12] Ionides, E. L., Breto, C. and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proc. National Academy of Sciences*, **103**, 18438-18443.

[13] Ionides, E.L., Bhadra, A., Atchadé, Y. and King, A.A. (2011). Iterated filtering. *Ann. Statist.*, **39**, 1776-1802.

[14] Kitagawa, G. and Sato, S. (2001). Monte Carlo smoothing and self-organising state-space model. In [8], 178–195. New York: Springer.

[15] Malik, S. and Pitt, M. K. (2011). Particle filters for continuous likelihood evaluation and maximisation. *J. Econometrics*, 165, 190-209.

[16] Olsson, J., Cappé, O., Douc, R. and Moulines, E. (2008). Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models. *Bernoulli*, **14**, 155-179.

[17] Poyiadjis, G., Doucet, A. and Singh, S.S. (2011). Particle approximations of the score and observed information matrix in state-space models with application to parameter estimation. *Biometrika*, **98**, 65-80.

[18] Whiteley, N. (2013) Stability properties of some particle filters. *Ann. Applied Proba.*, to appear.